# VAA-Weekly-Progress

09/18-09/24

# Context:

- Last week we discussed our Labeling tool that we would move forward with

- Worked towards establishing a definition of keypoints for our dataset

- Figured out formatting for loading data in order to train models

# Agenda :

- Discuss papers regarding labeling to understand intent of key point labeling in

  the respective papers

- Finalizing definition for key points by testing for accuracy amongst ourselves

  by using the videos extracted from last week

- New information regarding occluded key points

# Goals for the Week

- Read 300 faces and OpenApePose and document the papers
  - How and Why do they define their key points?
  - How could this paper be useful?
  - What does the paper not say?
- Watch videos on the selected images and observe if key points stay consistent in various frames.
- Have written definition for each key point to describe what landmarks to look out for/use in order to ideally place the key point.
  - Assigned to: Medha and Claire and Armaan
  - Check if AP-10k labels if the keypoint is visible or not. -> could use to parse AP-10k annotations to get only visible keypoint training data
- Read downstream uses for keypoints
  - APT-36k

# Papers

# OpenApePose - How to Format Our Training Dataset

Title: OpenApePose

Authors: Nisarg Desai, Praneet Bala, Rebecca Richardson, Jessica Raper, Jan Zimmerman, Benjamin Hayden

Publication year: 2023

Publication venue: Elife

What's new in this paper: Dataset of 71,868 annotated ape images across 6 different species (bonobos, chimpanzees, gorillas, orangutans, gibbons, and siamangs)

Data: OpenApePose is a set of ape images of four species from the Hominidae family: bonobos, chimpanzees, gorillas, orangutans, and several species from the Hylobatidae family, pooled into two categories of gibbons and siamangs. This dataset consists primarily of photographs taken at zoos, and also includes images from online sources, including publicly available photographs and videos.

Method: They used Hive AI to have annotators label 16 landmarks on each image(discussed more on the next slide), which then comprised a pose. They "ensured that the annotations were accurate by visualizing 5 random samples of 100 images with the annotations overlayed on the images, for each batch of 10,000 images, resulting in a total of ~2,500 inspected images."

# OpenApePose vs AP10k Keypoint Scheme
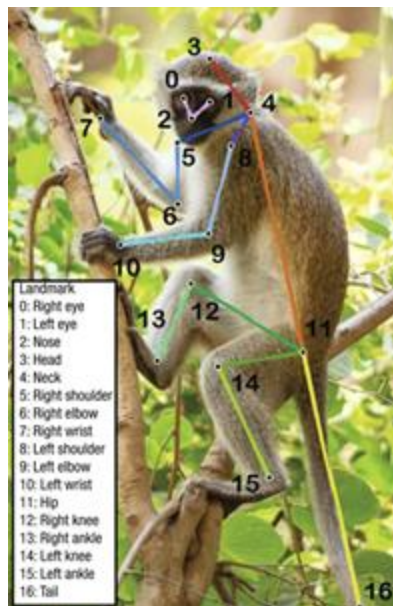


OpenMonkeyChallenge (2022)

(Source of OpenApePose scheme,
Tail keypoint gets excluded)
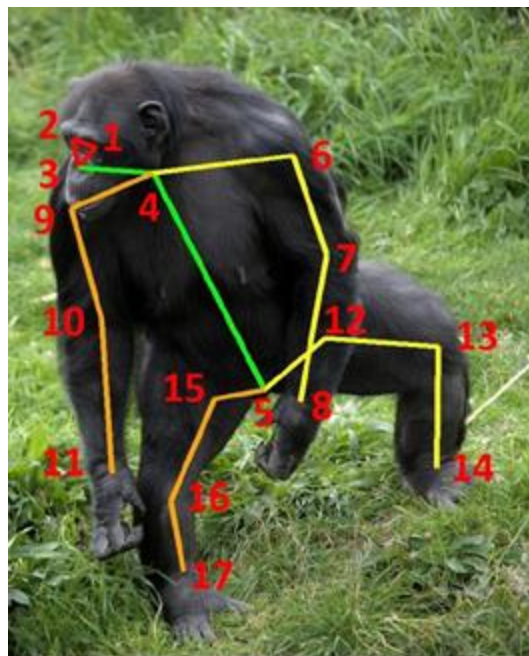


AP10k Github README

| Keypoint | Description | Keypoint | Description |
|---|---|---|---|
| 1 | Left Eye | 2 | Right Eye |
| 3 | Nose | 4 | Neck |
| 5 | Root of Tail | 6 | Left Shoulder |
| 7 | Left Elbow | 8 | Left Front Paw |
| 9 | Right Shoulder | 10 | Right Elbow |
| 11 | Right Front Paw | 12 | Left Hip |
| 13 | Left Knee | 14 | Left Back Paw |
| 15 | Right Hip | 16 | Right Knee |
| 17 | Right Back Paw | | |

- OpenApePose "Hip" keypoint maps to AP10k "Root of Tail"

- OpenApePose introduces a head keypoint

# Insights from OpenApePose

- Introduces evidence against generalizing keypoint labels across species

- Key major finding of the paper "large tailored species-specific annotated sets are superior to large multispecies sets"
  - Generalizing poses across different species limits a model's ability to learn species specific features

- OpenApePose model with a specific species removed from training data still performs better on that species than the OpenMonkeyPose model
  - (OpenMonkeyPose model trained on a much larger dataset that includes some Apes)

# Interesting Points Noted by the Authors

Main Points Authors Articulate:

- The OpenApePose data set keypoint definitions do not match AP-10k or COCO, but they claim and show HRNet-W48 trained on OpenApePose performs better on ape pose detection than either AP-10k or COCO trained on HRNet-W48
- They emphasized multiple times the need for large(~10,000s) and robust data sets on individual species or phylogentically close species as modern pose estimators do not generalize well. They emphasized it with one of their findings "that models with one ape species removed still perform better at tracking the held-out species more accurately than the OpenMonkeyPose model on that species."

Other Points of Notice:

- They also "manually sorted and cropped the images such that each cropped image contains the full body of at least one ape while minimizing repetitive poses to ensure a greater diversity of poses in the full dataset." And "We ensured that all cropped images have a resolution greater than or equal to 300×300 pixels."
- They also "manually removed images that were repetitive, for example in cases when the ape was at rest and did not change the pose and in cases when extracting frames from videos. We also removed images that were too blurry to identify the body landmarks for human annotators. "

Full OpenApePose Notes

# 300Faces

Title: 300 Faces In-The-Wild Challenge: database and results

Authors: C. Sagonas, E. Antonakos, G, Tzimiropoulos, S. Zafeiriou, M. Pantic

Publication year: 2016

Publication venue: Image and Vision Computing

What's new in this paper: The paper talks about the development of a semi-automatic annotation tool based on AOMs to standardize and reannotate the existing datasets.

Model architecture: Active Orientation Models (AOMs)

Data: Multi-PIE, XM2VTS, FRGC-V2, AR, LFPW, HELEN, AFW, and IBUG

Method: The paper introduces a semi-automatic annotation method using AOMs to re-annotate existing databases with a unified landmark configuration (68-point mark-up, since they differed)

# Insights from 300Faces

- Authors decided to move forward with the 68 key-point scheme used by Multi-PIE, since it is already the most popular and has the most amount of labeled data
    - Allowed for separation of a subset of 51 key-points, removing 17 key-points on the chin
    - These 51 key-points have much less variance than the points on the chin
    - "standard deviation is high for landmarks that do not have a clear semantic meaning"



C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, M. Pantic / Image and Vision Computing 47 (2016) 3–18
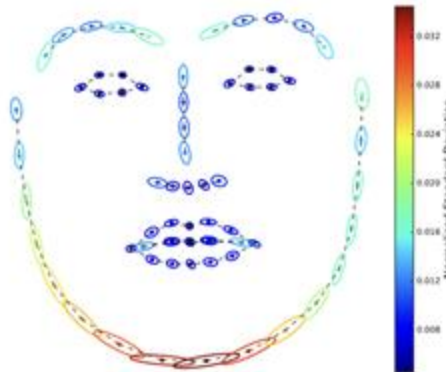
# Citations

**OpenApePose:**

@article{desai2023openapepose,
  title={OpenApePose, a database of annotated ape photographs for pose estimation},
  author={Desai, Nisarg and Bala, Praneet and Richardson, Rebecca and Raper, Jessica and Zimmermann, Jan and Hayden, Benjamin},
  journal={Elife},
  volume={12},
  pages={RP86873},
  year={2023},
  publisher={eLife Sciences Publications Limited}
}

**300 Faces:**

@article{sagonas2016300,
  title={300 faces in-the-wild challenge: Database and results},
  author={Sagonas, Christos and Antonakos, Epameinondas and Tzimiropoulos, Georgios and Zafeiriou, Stefanos and Pantic, Maja},
  journal={Image and vision computing},
  volume={47},
  pages={3--18},
  year={2016},
  publisher={Elsevier}
}

# Keypoint Definitions
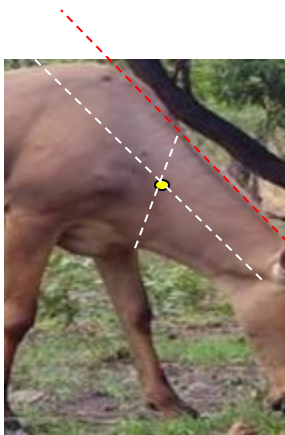
# Keypoint Definitions

- We created a formal definition of our keypoints: [Keypoint Definitions](Keypoint Definitions)

- To test out the effectiveness of the definitions, we gave the document to the other half of the team, and asked them to annotate frames from the videos

- LabelMe has tools that allow us to check the differences in annotation across annotators, which will allow us to see which keypoints have the largest discrepancies

- Includes section on how exactly to handle occlusions
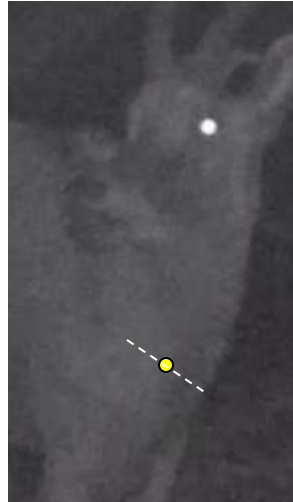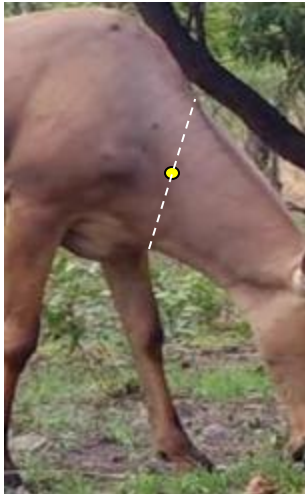
# Neck Keypoint

Original Definition
- For a side view: Place it at the center of the base of the side of the neck. Find the cross-section of the line coming where the skull connects to the neck (parallel to the spine) and the line at the base of the neck, where the neck connects to the torso
- For the front view: Place it at the center of the base of the front of the neck. Find the cross-section of the line coming where the skull connects to the neck (parallel to the spine) and the line at the base of the neck, where the neck connects to the torso
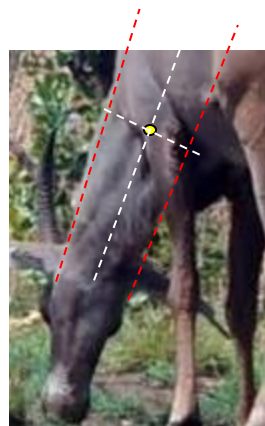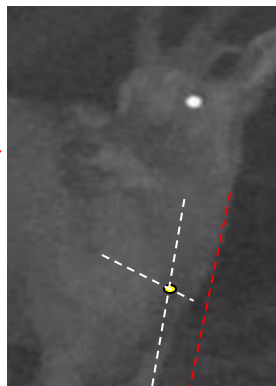
# Neck Keypoint

New Definition
- For a side view: Place it at the center of the base of the side of the neck. Find the line at the base of the neck, where the neck connects to the torso, and place it in the midpoint
- For the front view: Place it at the center of the base of the front of the neck. Find the line at the base of the neck, where the neck connects to the torso, and place it in the midpoint

# AP-10k Labeling of Occluded Key Points

Sample AP-10K keypoint annotations: "keypoints": [185, 173, 2, 119, 161, 2, 141, 195, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]}"

"For the occluded keypoints, they were asked to annotate keypoints whose location they could estimate based on body plan, pose, and the symmetry property of the body, where the length of occluded limbs or the location of occluded keypoints could be inferred from the visible limbs or keypoints. Other keypoints were left unlabeled."

No concrete explanation of the meaning of 2 and 0 in the keypoint annotation, but educated guess is that it means the keypoint was labeled or is unlabeled.

**Conclusion:** Cannot split AP-10k dataset into sets with all occluded keypoints labeled and all occluded keypoints unlabeled.

# Trying out the labeling definitions

# Next Steps

- Fully refine the keypoint definition based on the annotations from the team

- Begin labeling keypoints on the antelope data

- Begin training RTMPose on AP-10k to have a benchmark on if our training process is working in comparison to the pretrained model from OpenMMPose

# Personal Progress

# Armaan

- Annotated 100 images with labels (Set 5)
- Wrote a detailed method of handling occlusions. Visible in the Keypoint definition document.

# Shaan

- Annotated 100 images
- Read 300Faces and OpenApePose

# Aryan

- Annotated 100 images with labels (Set 0)
- Discussed and formulated formal definition of keypoints with team

# Medha

- Annotated 100 images with labels (Set 4)
- Came up with a formal definition of the eyes, nose, neck, tail, and paw keypoints

# Claire

- Annotated 100 images with labels (Set 2)
- Came up with a formal definition of the hips, knees, shoulders, and elbows keypoints

# Parth

- Annotated 100 images (set 6)
- Briefly read 300Faces to identify any ideas pertaining to keypoint labeling
- Labeled 3 images from 1 video for trial on the definitions that we agreed upon